Review

Discrete RV's:

prob'y fctn: p(x) = Pr(X = x)

cdf: F(x) = Pr($X \leq$ x)

E(X) = $\sum_x$ x p(x)

SD(X) = $\sqrt{\text{E}\{(X - \text{E}\,X)^2\}}$

Binomial(n,p):

no. successes in n indep. trials where

Pr(success) = p in each trial

If $X \sim$ binomial(n,p), then:

Pr(X = x) = $\binom{n}{p} p^x (1-p)^{n-x}$

E(X) = $n\,p$

SD(X) = $\sqrt{np(1-p)}$

Binomial random variable

Number of successes in n trials where:

- Trials independent
- p = Pr(success) is constant

The number of successes in n trials does not necessarily follows a binomial distribution.

Deviations from the binomial:

- Varying p
- Clumping or repulsion (non-independence)

Examples

Consider Mendel's pea experiments. (Purple or white flowers; purple dominant to white.)

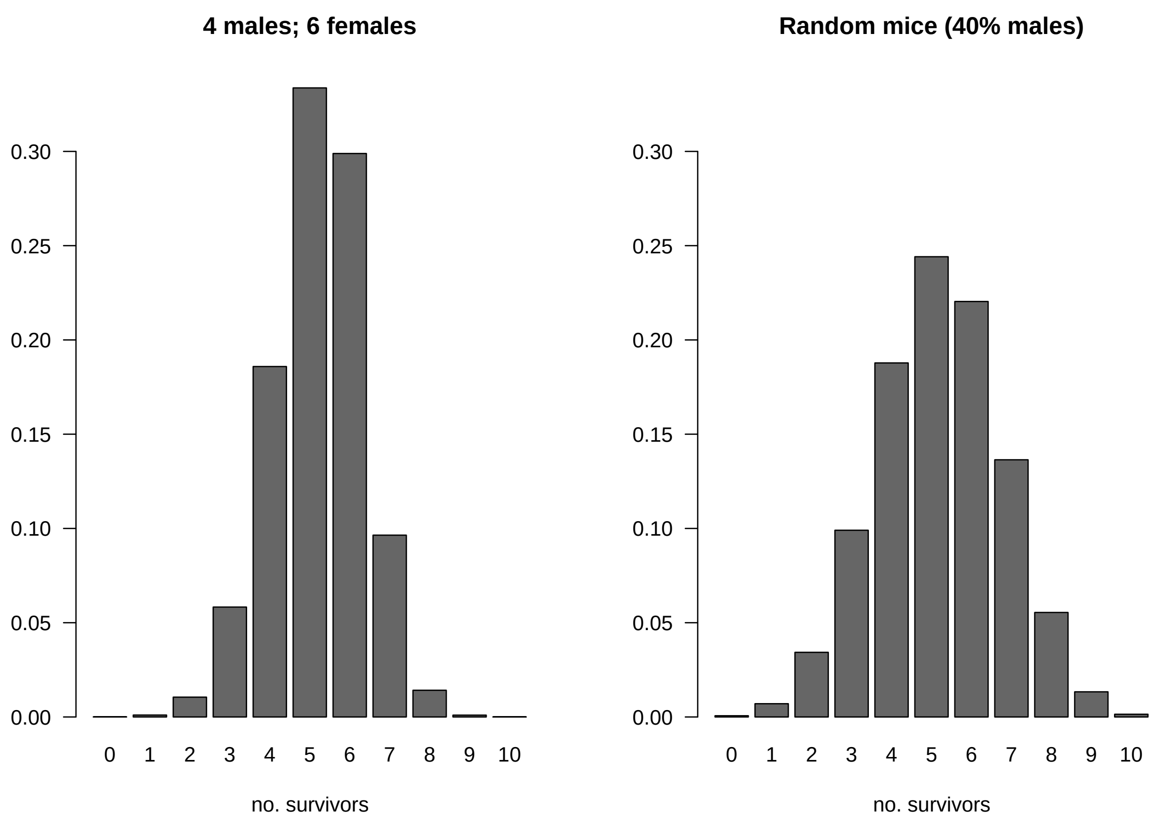
- Pick a random F_2. Self it and acquire 10 progeny. The number of progeny with purple flowers is *not* binomial (unless we condition on the genotype of the F_2 plant).

- Pick 10 random F_2's. Self each and take one child from each. The number of progeny with purple flowers *is* binomial. (p = (1/4) $\times$ 1 + (1/2) $\times$ (3/4) + (1/4) $\times$ 0 = 5/8.)

Suppose Pr(survive | male) = 10% but Pr(survive | female) = 80%.

- Pick 4 male mice and 6 female mice. The number of survivors is *not* binomial.

- Pick 10 random mice (with Pr(mouse is male) = 40%). The number of survivors *is* binomial.



Y = a + b X

Suppose X is a discrete random variable with probability function p, so that p(x) = Pr(X = x).

Expected value (mean): $E(X) = \sum_x x\, p(x)$

Standard deviation (SD): $SD(X) = \sqrt{\sum_x [x - E(X)]^2\, p(x)}$

Let $Y = a + b\,X$ where a and b are numbers. Then Y is a random variable (like X), and

$E(Y) = a + b\,E(X)$

$SD(Y) = |b|\,SD(X)$

In particular, if $\mu = E(X)$, $\sigma = SD(X)$, and $Z = (X - \mu) / \sigma$, then

$E(Z) = 0$ and $SD(Z) = 1$

Example

Suppose $X \sim \text{binomial}(n, p)$.
(The number of successes in n independent trials where p = Pr(success).)

Then $E(X) = n\,p$ and $SD(X) = \sqrt{n\,p\,(1-p)}$

Let $P = X / n$ = proportion of successes.

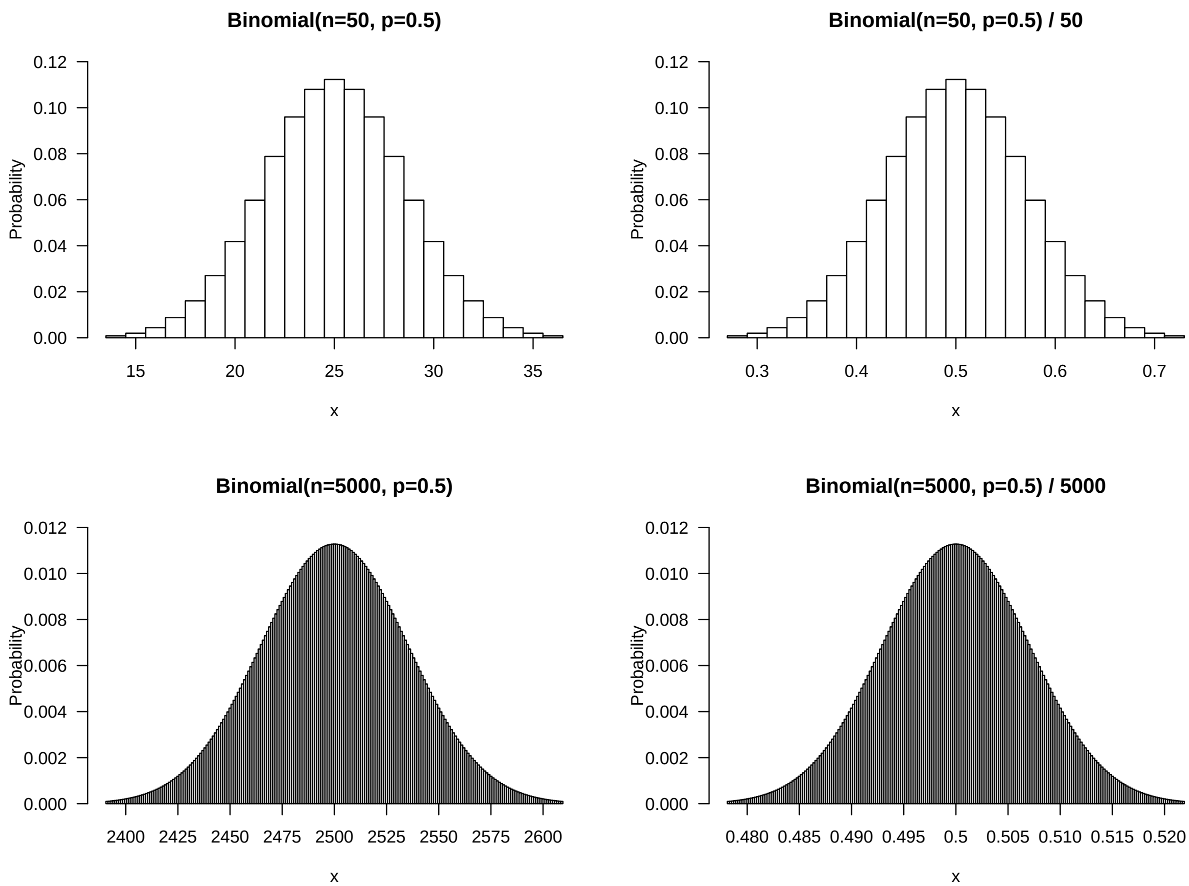
$E(P) = E(X / n) = E(X) / n = p$.

$SD(P) = SD(X / n) = SD(X) / n = \ldots = \sqrt{p\,(1-p)/n}$

Toss a fair coin n times and count X = number of heads and $P = X/n$.

For n=50: $E(X) = 25$, $SD(X) \approx 3.5$, $E(P) = 0.5$, $SD(P) \approx 0.07$
$\Pr(X = 25) = \Pr(P = 0.5) \approx 0.11$

For n=5000: $E(X) = 2500$, $SD(X) \approx 35$, $E(P) = 0.5$, $SD(P) \approx 0.007$
$\Pr(X = 2500) = \Pr(P = 0.5) \approx 0.011$



Poisson distribution

Consider a binomial(n, p) where

- n is really large
- p is really small

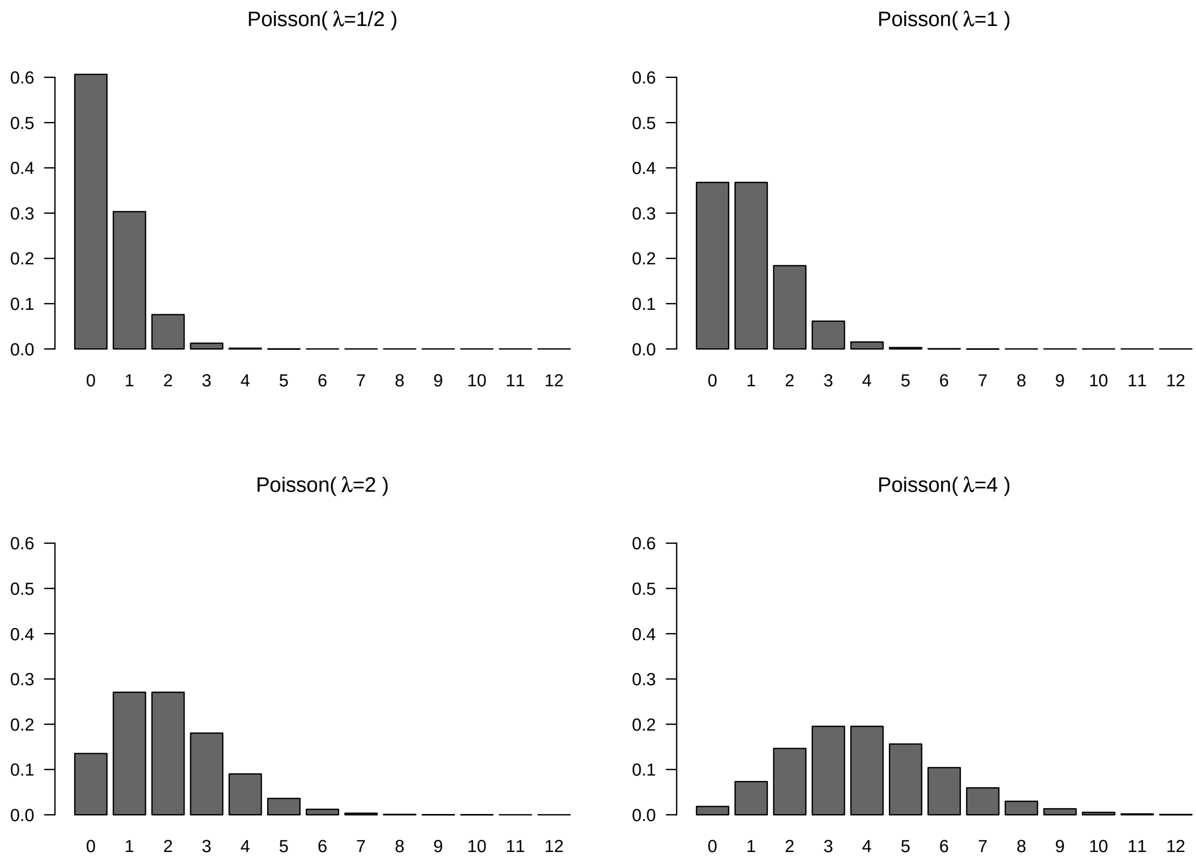
For example, suppose each well in a microtiter plate contains 50,000 T cells, and that 1/100,000 cells respond to a particular antigen.

Let X be the number of responding cells in a well.

In this case, X follows a Poisson distribution.

Let $\lambda = n \times p = \mathrm{E}(X)$. Then $\mathrm{p(x)} = \Pr(X = \mathrm{x}) = \mathrm{e}^{-\lambda}\lambda^{\mathrm{x}}/\mathrm{x}!$

Note that $\mathrm{SD(X)} = \sqrt{\lambda}$.



Example

Suppose there are 100,000 T cells in each well of a microtiter plate. Suppose that 1/80,000 T cells respond to a particular antigen.

Let X = number of responding T cells in a well.

$X \sim$ Poisson(λ = 1.25).

E(X) = 1.25; SD(X) = $\sqrt{1.25} \approx$ 1.12.

Pr(X = 0) = exp(–1.25) $\approx$ 29%.

Pr($X > 0$) = 1 – exp(–1.25) $\approx$ 71%.

Pr(X = 2) = exp(–1.25) $(1.25)^2/2 \approx$ 22%.

In R

The following functions act just like `rbinom`, `dbinom`, etc., for the binomial distribution:

```
rpois(m, lambda)
dpois(x, lambda)
ppois(q, lambda)
qpois(p, lambda)
```


Continuous random variables

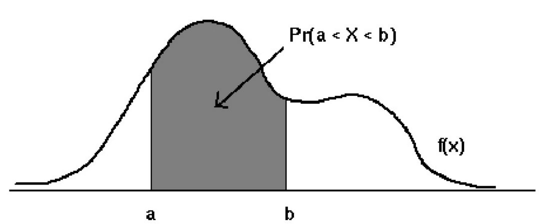
Suppose X is a continuous random variable.

Instead of a probability function, X has a probability density function (pdf), sometimes called just the density of X.

$f(x) \geq 0$

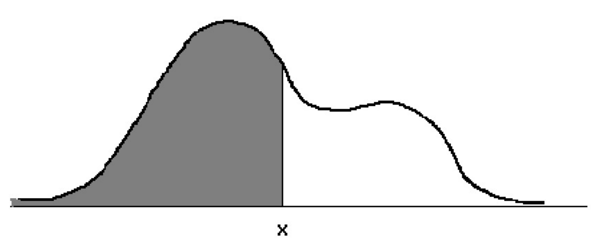
$\int_{-\infty}^{\infty} f(x)\,dx = 1$

Areas under curve = probabilities



Cumulative distr'n func'n (cdf):

$F(x) = \Pr(X \leq x) =$



Means and SDs

Expected value (mean):

Discrete RV: E(X) = $\sum_x$ x p(x)

Continuous RV: E(X) = $\int_{-\infty}^{\infty}$ x f(x) dx

Standard deviation (SD):

Discrete RV: SD(X) = $\sqrt{\sum_x [x - E(X)]^2 \, p(x)}$

Continuous RV: SD(X) = $\sqrt{\int_{-\infty}^{\infty} [x - E(X)]^2 \, f(x) \, dx}$

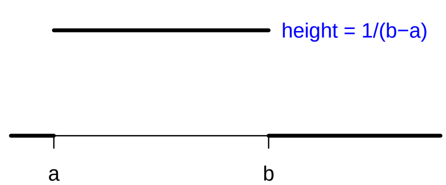
Example: Uniform distribution

$X \sim$ Uniform(a, b)

i.e., draw a number at random from the interval (a, b).

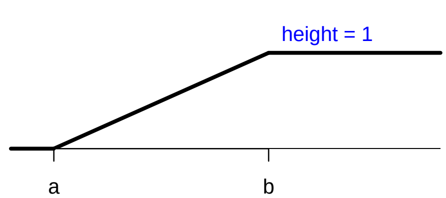
Density function:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$



E(X) = (b+a)/2 SD(X) = (b – a)/$\sqrt{12}$ $\approx$ 0.29 $\times$ (b – a)

Cumulative dist'n fdn (cdf):



The normal distribution

By far the most important distribution:

The Normal distribution (also called the Gaussian distribution)

If $X \sim \text{N}(\mu, \sigma)$, then

The pdf of X is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

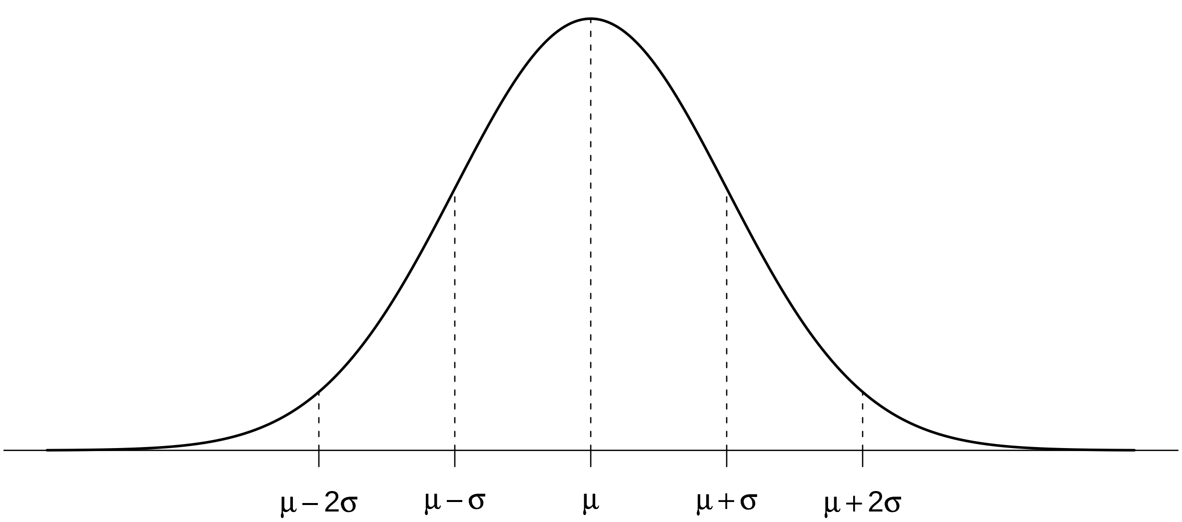
Also $E(X) = \mu$ and $SD(X) = \sigma$.

Of great importance: If $X \sim \text{N}(\mu,\sigma)$ and $Z = (X - \mu) / \sigma$,

Then $Z \sim \text{N}(0, 1)$.

This is the "Standard normal distribution".

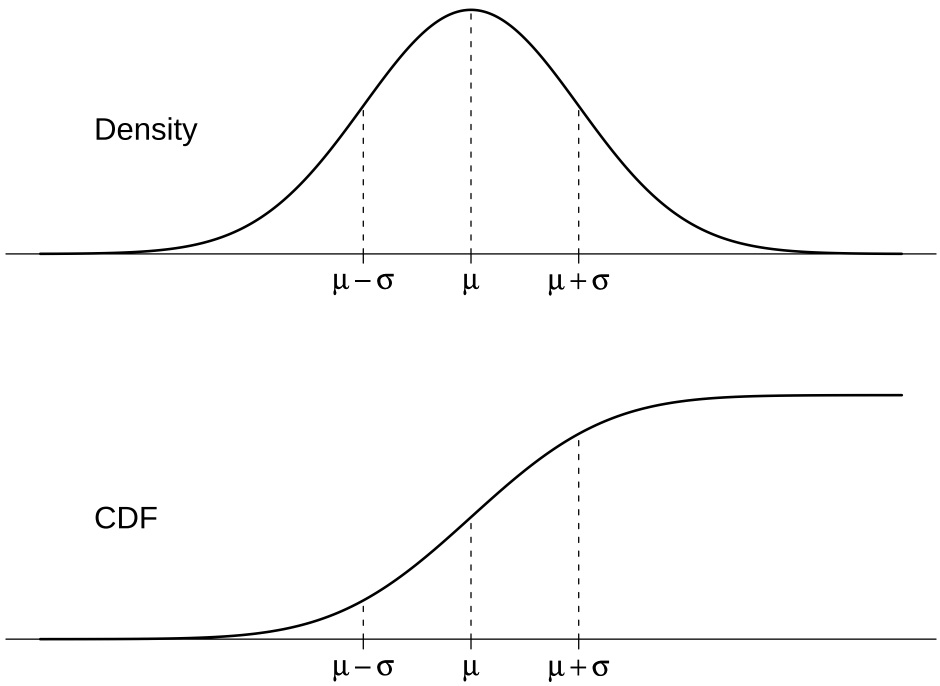
The normal distribution



$$\Pr(\mu - \sigma \le X \le \mu + \sigma) \approx 68\%$$

$$\Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 95\%$$

The normal CDF



Calculations with the normal curve

In R:
- Convert to a statement involving the cdf
- Use the function `pnorm`

  (See also `rnorm`, `dnorm`, and `qnorm`.)

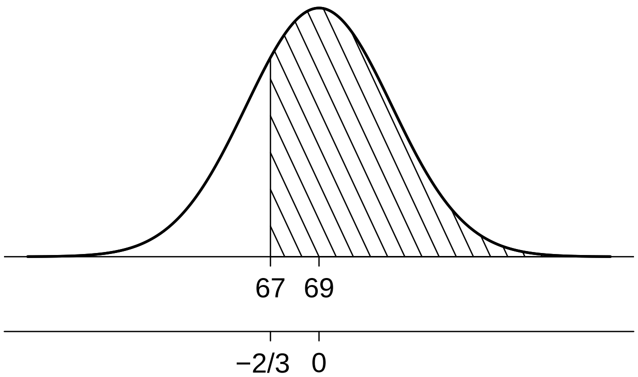
With a table:
- Convert to a statement involving the standard normal
- Convert to a statement involving the tabulated areas
- Look up the values in the table

Draw a picture!

Examples

Suppose the heights of adult males in the U.S. are approximately normal distributed, with mean = 69 in and SD = 3 in.

What proportion of men are taller than 5'7"?

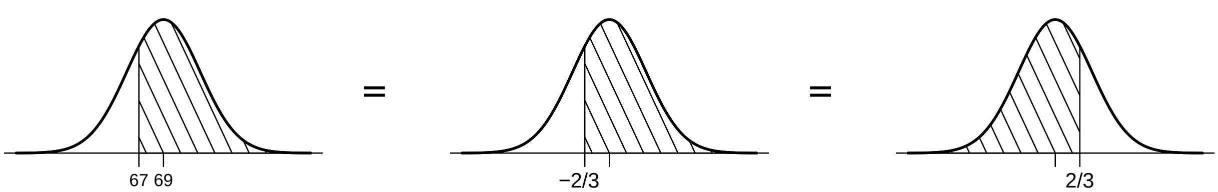


$X \sim \text{N}(\mu{=}69, \sigma{=}3)$

$Z = (X - 69)/3 \sim \text{N}(0,1)$

$\Pr(X \geq 67) = \Pr(Z \geq (67 - 69)/3)$
$= \Pr(Z \geq -2/3)$

R (or a table)

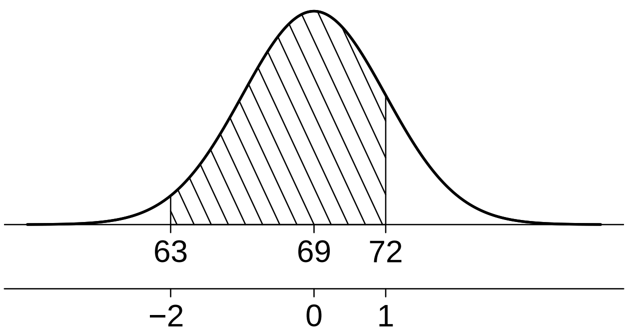


Use either `pnorm(2/3)` or `1 - pnorm(67, 69, 3)` or `pnorm(67, 69, 3, lower.tail=FALSE)`

The answer: 75%.

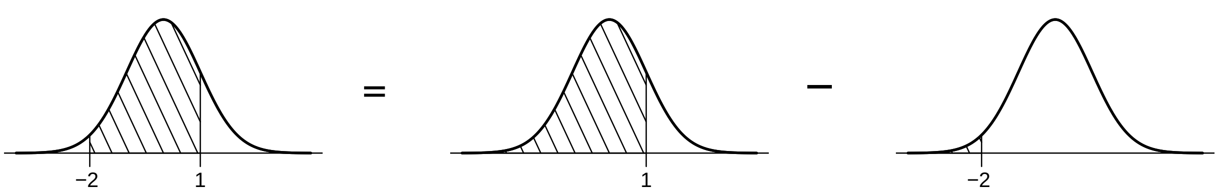
Another calculation

What proportion of men are between 5'3" and 6'?



$\Pr(63 \leq X \leq 72) = \Pr(-2 \leq Z \leq 1)$

R (or a table)




```
pnorm(72, 69, 3) - pnorm(63, 69, 3)
```

or

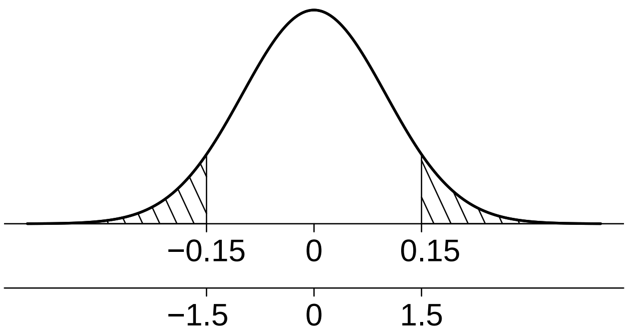
```
pnorm(1) - pnorm(-2)
```

The answer: 82%.

One last example

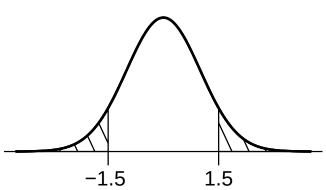
Suppose that the measurement error in a laboratory scale follows a normal distribution with mean = 0 mg and SD = 0.1 mg.

What is the chance that the absolute error in a single measurement will be greater than 0.15 mg?

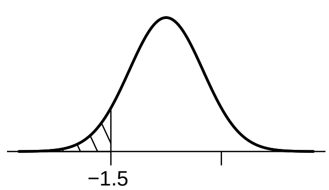


$\Pr(|X| \geq 0.15) = \Pr(|Z| \geq 1.5)$

R (or a table)



$= \quad 2 \times$




```
2 * pnorm(-0.15, 0, 0.1)
```

or

```
2 * pnorm(-1.5)
```

The answer: 13%.